

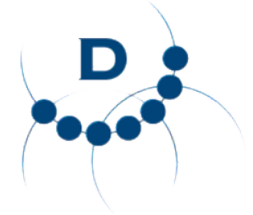


Service Centers as backbone of a D-SPIN infrastructure - requirements and current state -

Peter Wittenburg
MPI for Psycholinguistics



- technical infrastructure
- required pillars
- centres network as backbone
- requirements for centres
- state centres network
- state pillars
- relations to e-Infrastructures
- summary



J. Taylor:

“eScience is about global collaboration in key areas of science and the **next generation of infrastructures** that will enable it”

Need to build new types of platforms

- to allow researchers to combine existing resources and tools easily to new ones to tackle the **big challenges**
- to increase the **productivity** of all interested researchers, since currently too much time is wasted by preparatory work

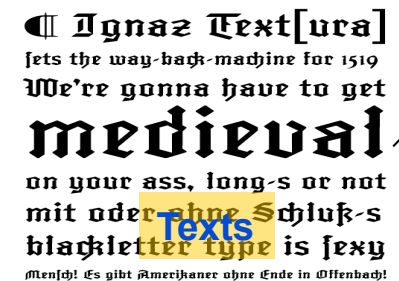
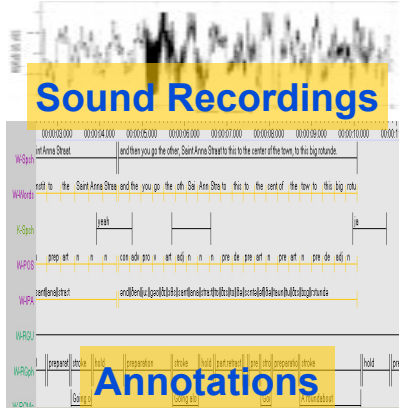


Like to compare it with the tracks and signalling for the ICE

Current Download-First Paradigm

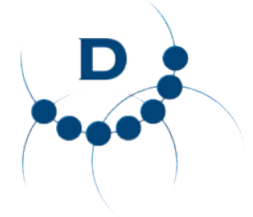
D-SPIN

CLARIN



- installing
- downloading
- adapting
- converting
- scripting
- etc etc

- some like it very much as I liked the VW Käfer
- not very efficient, rights problems, etc
- many are cut off since one needs IT skills
- cyberspace needs to overcome this scenario

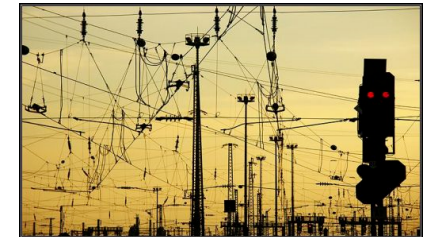


- which kind of pillars will make up the required infrastructure suitable for e-Research
- need to solve the integration and interoperability issues
 - can we rely on the availability of LRT?
 - > stable repositories
 - > stable service execution frameworks
 - can we see the resources, tools and services?
 - > joint metadata domain
 - can we guarantee stability of references?
 - > persistent identifier service



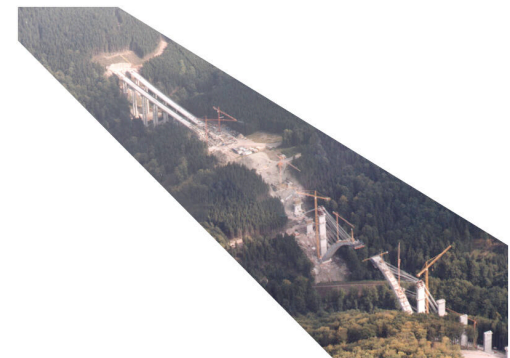


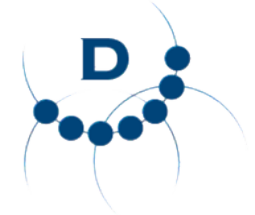
- need to solve the integration and interoperability issues
 - can we allow virtual collection & workflow building?
 - > domain of trust
 - > distributed authentication&authorization
 - can we trust our servers?
 - > domain of certificates
 - can we find things?
 - > need many registries essential for a distributed arch
 - **can we support easy crosswalk?**
 - > pivot formats and converters
 - > solve the semantic gap -> ISOcat, etc ?
 - > build framework for a service oriented architecture



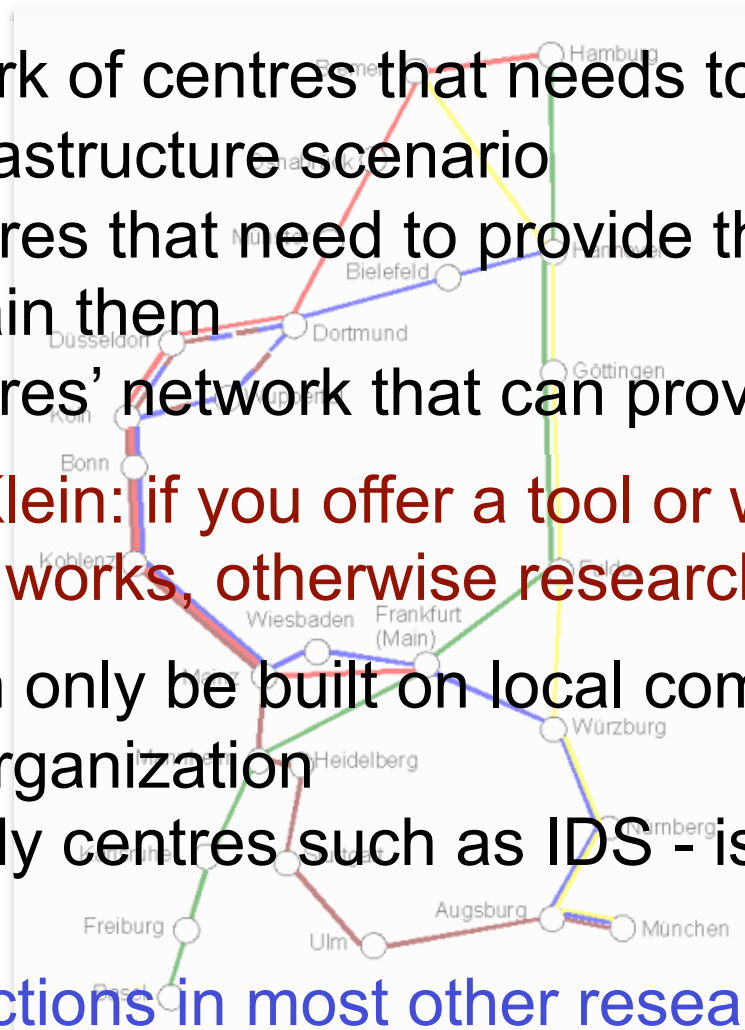


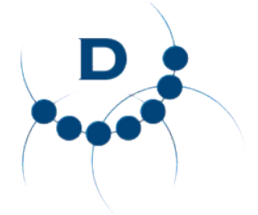
- building an infrastructure is expensive and needs time
- building the “ICE trains” is not infrastructure
- building Virtual Research Environments, tools and services is a task of research
- creating new resources dito
- lifetime infrastructure >> lifetime of tools/services
- essential questions
 - how can we make investments persistent?
 - how can we guarantee stability and availability of all services?
 - who will maintain all this?





- it's a network of centres that needs to act as backbone in a cyberinfrastructure scenario
- it's the centres that need to provide the essential services and maintain them
- it's the centres' network that can provide redundancy
- **Wolfgang Klein: if you offer a tool or whatsoever, take care that it works, otherwise researchers will not use it!!**
- centres can only be built on local commitments of the research organization
- have already centres such as IDS - is something new?
- similar directions in most other research infrastructures





- centres need
 - to be devoted to continuous service optimization without own agenda
 - to offer services that can be integrated into the researchers' workflow, i.e.
 - quick search and easy execution
 - frequently random access patterns
 - unrestricted access (not just a restricted interface)
 - easy annotations in workspaces
 - seamless transition between local and infra resources where possible (export into spreadsheet etc)
 - etc (Hinrichs -> coming Prague Workshop)

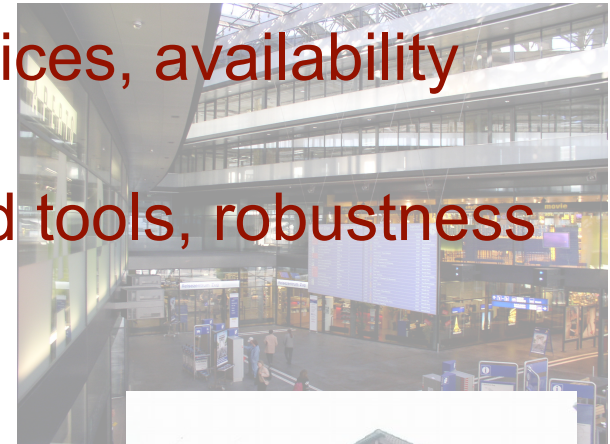
Types of Centres

D-SPIN

CLARIN



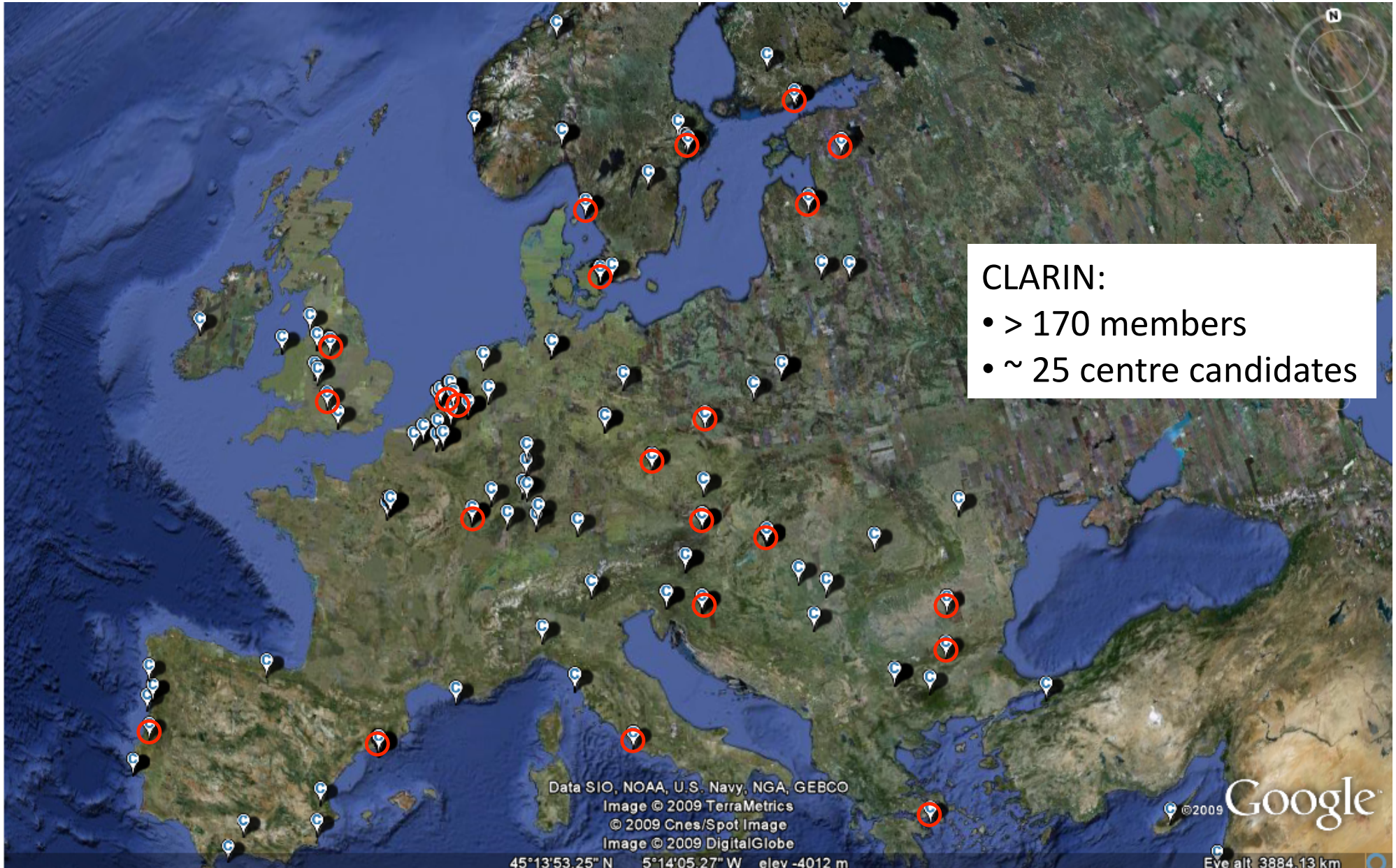
- Infrastructure centres sensitive services, availability
- LRT service centres resources and tools, robustness
- Metadata providers just metadata
- Respected providers just web portal
- external service providers (D-SPIN/CLARIN does not have to offer all types of infrastructure services -> later)



CLARIN/D-SPIN Centres

D-SPIN

CLARIN



CLARIN:

- > 170 members
- ~ 25 centre candidates

Data SIO, NOAA, U.S. Navy, NGA, GEBCO
Image © 2009 TerraMetrics
© 2009 Cnes/Spot Image
Image © 2009 DigitalGlobe
45°13'53.25" N 5°14'05.27" W elev -4012 m

©2009 Google

Eye alt 3884.13 km

Requirements for Centres

D-SPIN

CLARIN



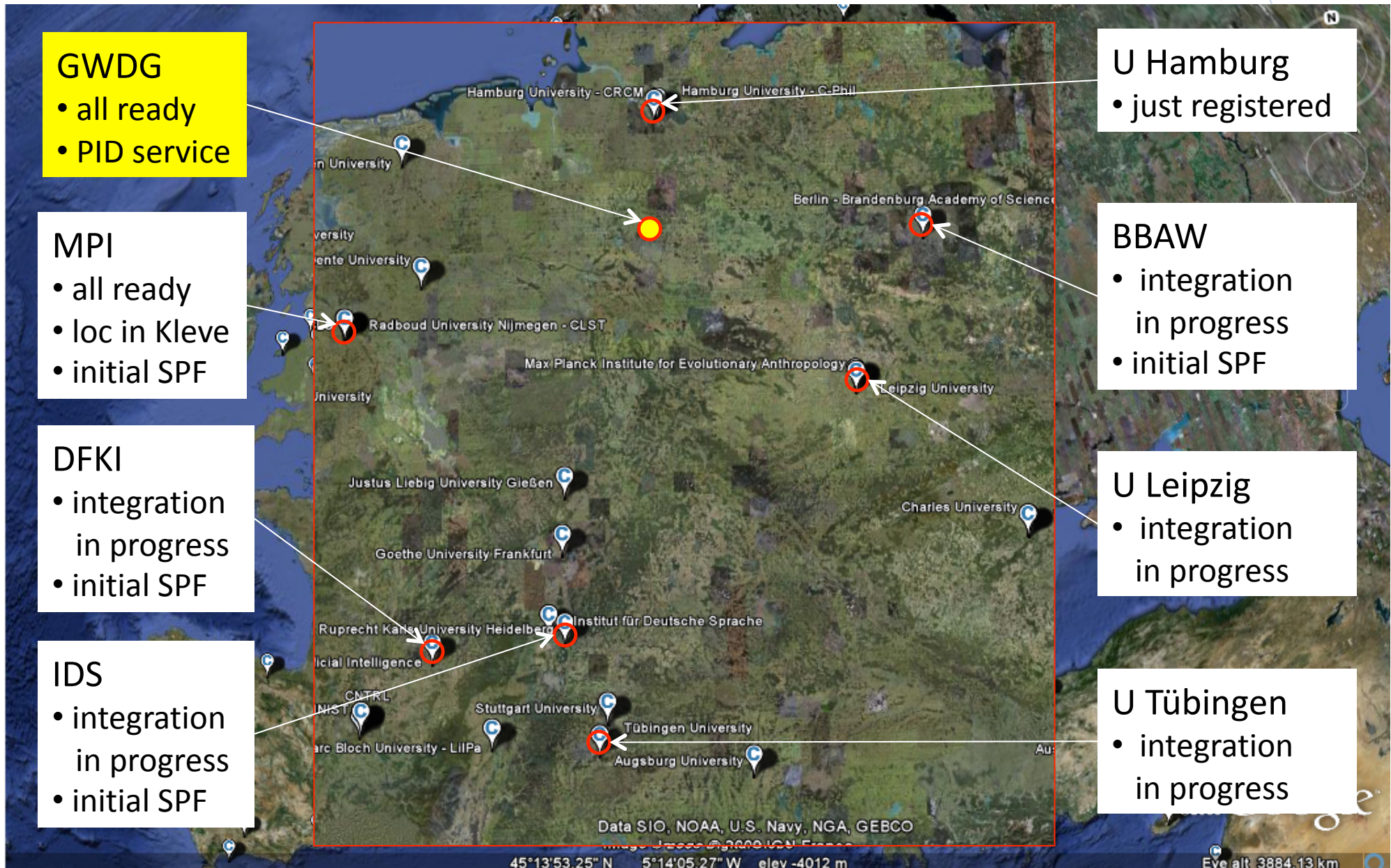
- offer all required attributes for centre registry -> self/D-SPIN
- certify all your servers -> DFN
- integrate to national identity federation -> DFN AAI
- integrate to LRT service provider federation -> D-SPIN
- setup a proper repository -> self /own costs
- associate PIDs with your resources (services) -> self
- create and offer (OAI PMH) compliant metadata -> self
- make your offers explicit -> self
- carry out quality self assessments -> self/monitored by DSA
- allow service monitoring -> D-Grid/D-SPIN instance
- offer workspaces/execution spaces -> self/PARADE
- take care of archiving solution -> MPI/?/PARADE
- etc

State Centres Network

D-SPIN



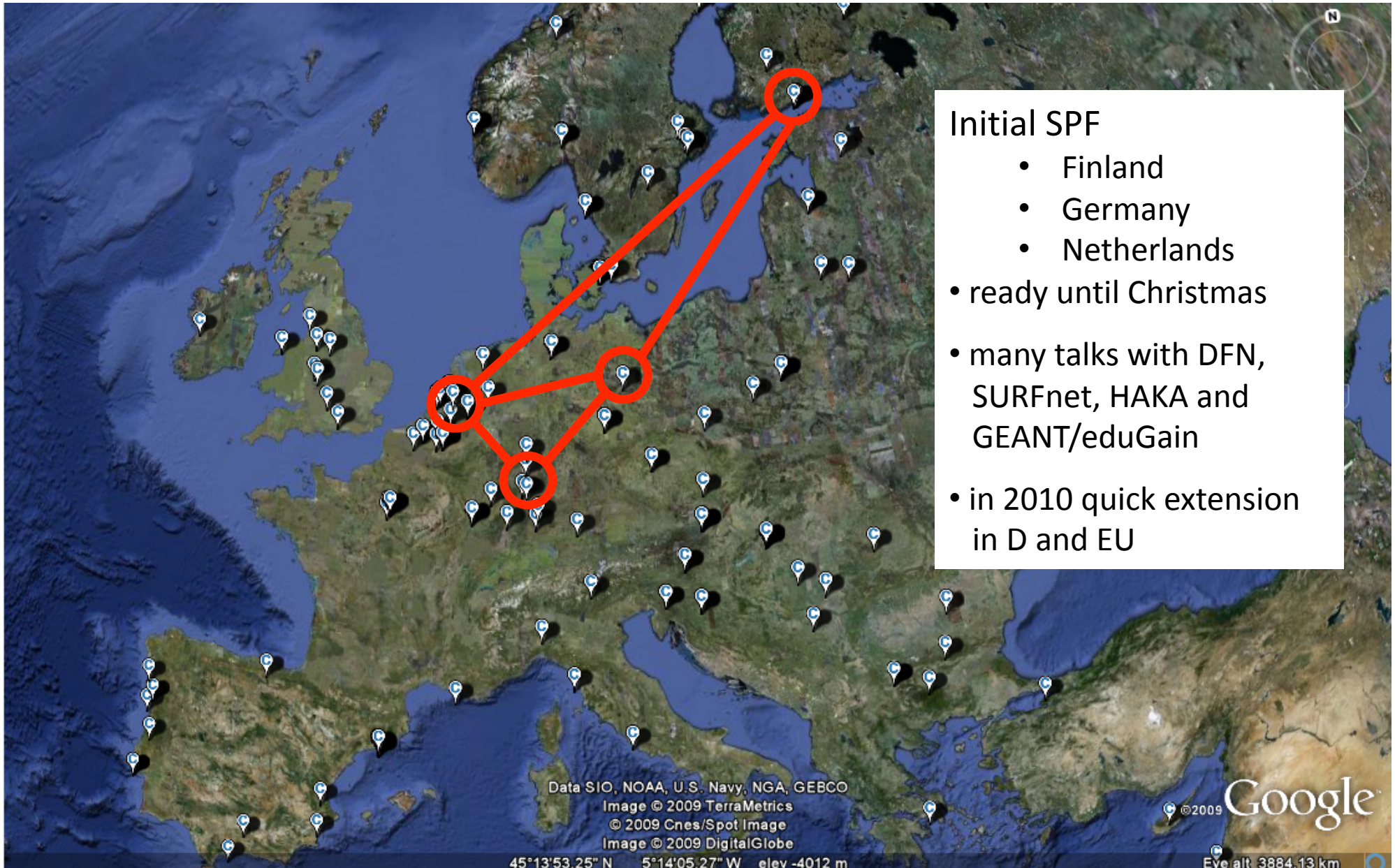
CLARIN



Initial Federation - State AAI

D-SPIN

CLARIN



Initial SPF

- Finland
- Germany
- Netherlands
- ready until Christmas
- many talks with DFN, SURFnet, HAKA and GEANT/eduGain
- in 2010 quick extension in D and EU



- GWDG offers stable PID service based on handles
- most robust system for registration & resolution
- REST API is ready and can be used
- open for all CLARIN “centres” and others (SUB, MPG, ...)
(makes only sense for centres that have stable repository)
- recently extended at European level for redundancy

- no one in D-SPIN obliged to use that service but ...

Munich, October 17 2009

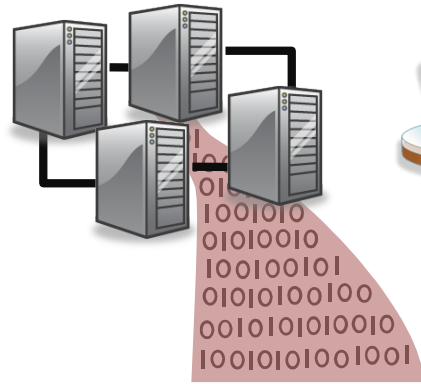


State Pillars - PIDs

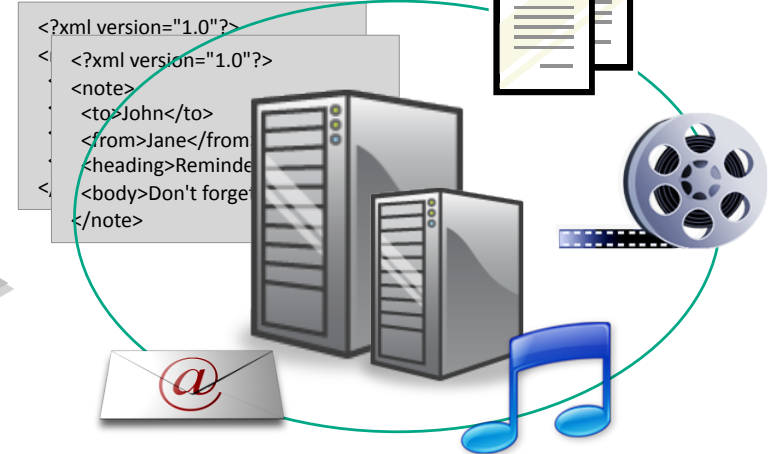
D-SPIN

D

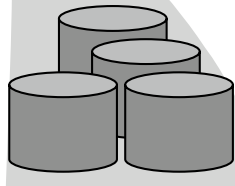
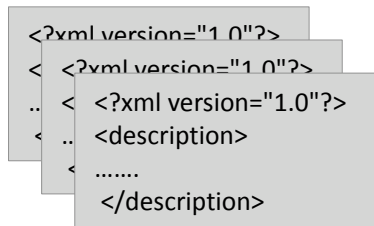
CLARIN



Client

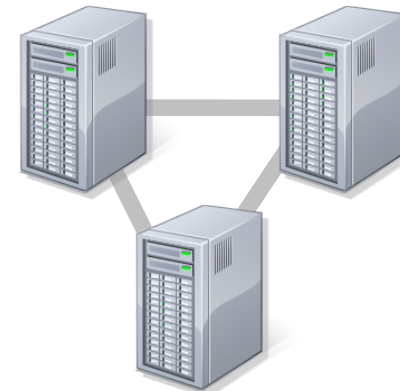


Repositories / Collections



Resource Discovery

Search Engines, Metadata Databases, Catalogues, Guides, etc.
Munich, October 17 2009

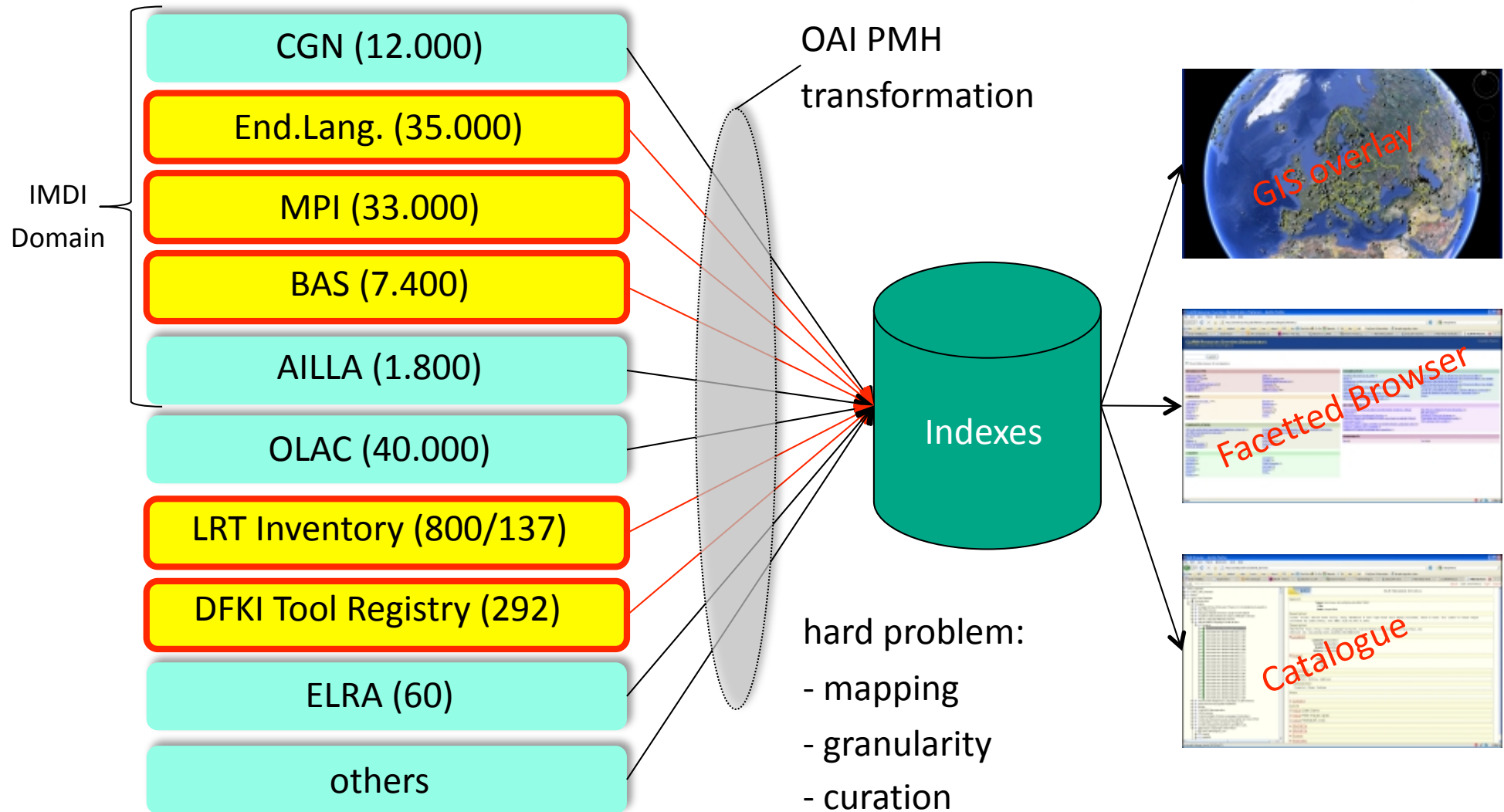


Identifier Resolution System

State Pillars - Metadata Domain

D-SPIN

CLARIN



State Pillars - Metadata Domain

D-SPIN

CLARIN



CLARIN Virtual Language Observatory - Resources

Powered by Flamenco

Demonstrator with IMDI, OLAC, ELRA and CLARIN data (contact: vlv@clarin.eu)

COUNTRY: Germany

Subcategories in **ORGANISATION**:

[Bavarian Archive for Speech Signals \(BAS\)](#) (7417)

[Berlin-Brandenburgische Akademie der Wissenschaft,](#)

[Digitales Wörterbuch der deutschen Sprache des 2](#) (1)

[Berlin-Brandenburgische Akademie der Wissenschaften](#) (1)

[Berlin-Brandenburgische Akademie der Wissenschaften, Das digitale Wörterbuch der deutschen Sprache](#) (2)

[Berlin-Brandenburgische Akademie der Wissenschaften, Das digitale Wörterbuch der deutschen Sprache](#) (1)

[Chemnitz University of Technology](#) (1)

[Das digitale Wörterbuch der deutschen Sprache des 19. Jahrhunderts, Berlin-Brandenburgische Akademie der Wissenschaften](#) (1)

[Freie Universität Berlin](#) (1391)

[German Research Foundation](#) (1)

[Hochschule Magdeburg](#) (10)

[Humanethologisches Filmarchiv der Göttinger Gesellschaft und Humanwissenschaftliches Zentrum für Kognitionswissenschaften](#) (1)

[IDS Mannheim](#) (11)

[Institut fuer Deutsche Sprache](#) (1)

[Institute of Indology and Tamil Studies, Cologne University](#) (1)

[Kompetenzzentrum für elektronische Verzeichnungs- und Publikationsverfahren](#) (1)

[MPI for Human History](#) (1)

[MPI für Bildungsforschung](#) (1)

[Magdeburgische Technische Universität, Applied Sciences](#) (30)

[Magdeburgische Technische Universität, Applied Sciences](#) (1166)

[Max Planck Institute for Advanced Study \(NIAS\), Collegium](#) (1)

[Max Planck Institute for Advanced Study \(NIAS\), Collegium](#) (4)

[Max Planck Institute for Advanced Study \(NIAS\), Collegium](#) (1)

[Max Planck Institute for Advanced Study \(NIAS\), Collegium](#) (1)

[Max Planck Institute for Advanced Study \(NIAS\), Collegium](#) (1)

[Max Planck Institute for Advanced Study \(NIAS\), Collegium](#) (1)

[Max Planck Institute for Advanced Study \(NIAS\), Collegium](#) (1)

[University of Konstanz](#) (1)

[University of Leipzig • Department of Computer Science •](#)

[NLP Group](#) (1)

[University of Surrey](#) (2)

[University of Surrey, Surrey Morphology Group](#) (1)

strong German contribution
who is responsible for quality and curation

Munich, October 17 2009

much curation still to be done

D-SPIN Language Resources

D-SPIN



CLARIN

many resources from D-SPIN colleagues and beyond

many German resources

much curation needed

- A -
[Abkhazian](#) (1)
[Afrikaans](#) (1)
[Albanian](#) (3)
[Arabic](#) (1)
[Armenian](#) (1)
[Avestan](#) (1)

- B -
[Bulgarian](#) (2)

- C -
[Catalan](#) (1)
[Chinese](#) (1)
[Church Slavonic](#) (2)
[Croatian](#) (2)
[Czech](#) (3)

- D -
[Danish](#) (2)
[Deutsche Gebärdensprache](#) (9)
[Divehi](#) (1)
[Dutch](#) (3)

- E -
[English](#) (347)
[Esperanto](#) (1)
[Estonian](#) (2)

- F -
[Finnish](#) (1)
[French](#) (55)

- G -
[Georgian](#) (3)
[German](#) (13198)
[German Sign Language](#) (777)
[Greek](#) (2)

- H -
[Hebrew](#) (1)
[Hindi](#) (1)
[Hungarian](#) (2)

- I -

- K -
[Korean](#) (1)

- L -
[Latin](#) (2)
[Latvian](#) (2)
[Lithuanian](#) (3)
[Lower Sorbian](#) (1)

- M -
[Malay](#) (1)

- N -
[Norwegian](#) (2)

- O -
[Occitan](#) (1)
[Ossetian](#) (1)

- P -
[Persian](#) (3)
[Polish](#) (609)
[Portuguese](#) (2)

- R -
[Romanian](#) (2)
[Russian](#) (33)

- S -
[Sanskrit](#) (4)
[Serbian](#) (1)
[Serbo-Croatian](#) (4)
[Slovak](#) (2)
[Slovenian](#) (3)
[Spanish](#) (1)
[Sundanese](#) (1)
[Swedish](#) (3)

- T -
[Tagalog](#) (1)
[Turkish](#) (9)

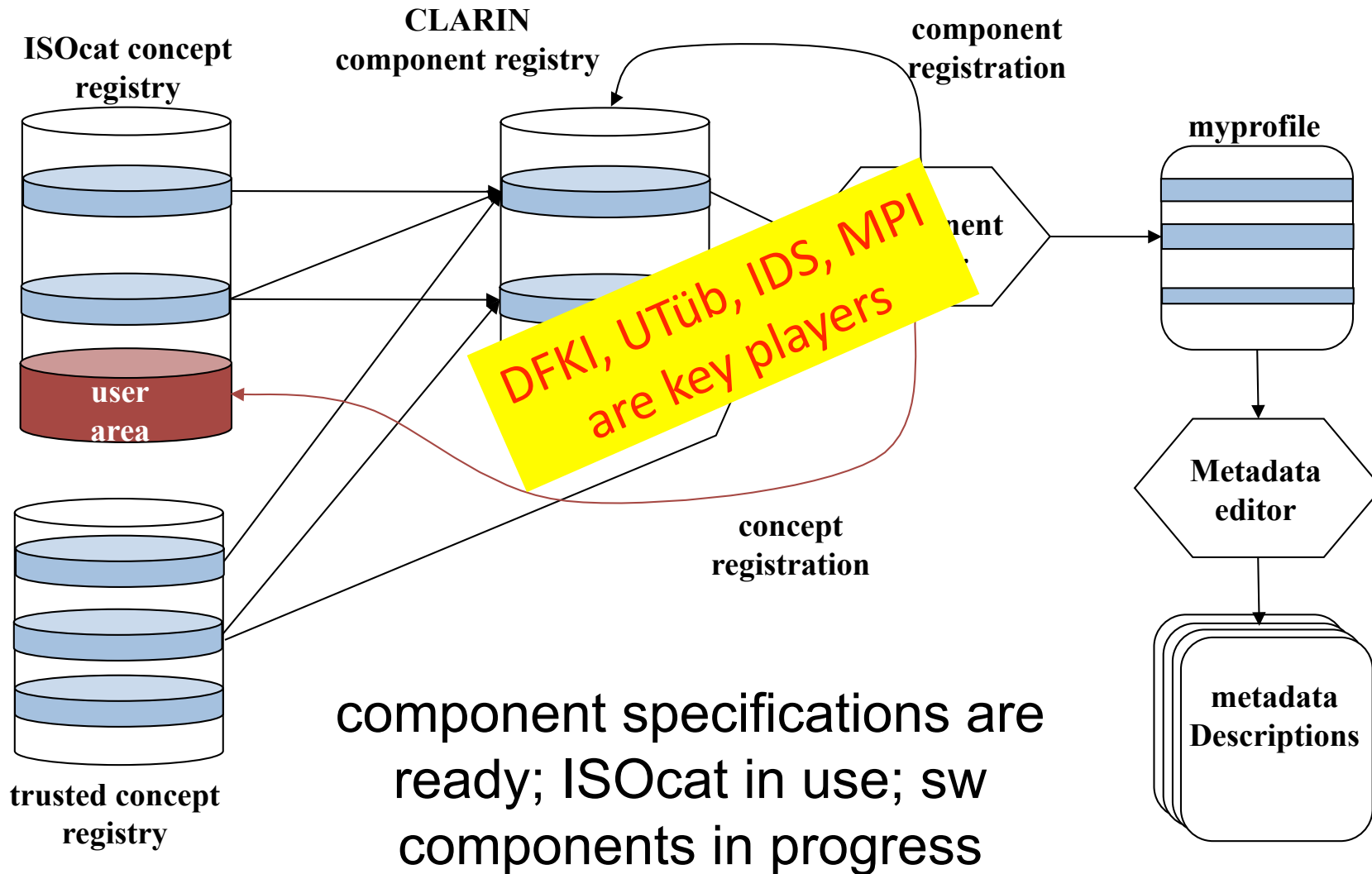
- U -
[Ukrainian](#) (1)
[Upper Sorbian](#) (1)

Munich, October 17 2009

Component Metadata Framework

D-SPIN

CLARIN



ISOcat Data Category Registry

D-SPIN

CLARIN



ISOcat - Data Category Registry - Mozilla Firefox

ISOcat - Web interface - Mozilla Firefox

ISOcat

enter keywords here

#	Name	Version	Administration st	Registration stabi	Check	Type	Owned by	Scope
2505	Address	0.0.0	private	candidate	⚠	open	Wittenburg, Pete	public
2562	Annotation Format	0.0.0	private	candidate	⚠	open	Wittenburg, Pete	public
2462	Annotation Level Typ	0.0.0	private	candidate	⚠	open	Wittenburg, Pete	public
2507	Annotation Stand-off	0.0.0	private	candidate	⚠	closed	Wittenburg, Pete	public
2508	Annotation Workflow	0.0.0	private	candidate	⚠	open	Wittenburg, Pete	public
2548	Anonymization Fl	0.0.0	private	candidate	⚠	closed	Wittenburg, Pete	public
2623	Anthron	0.0.0	private	candidate	⚠	simple	Wittenburg, Pete	public

1. Administration Information Section

record

DFKI, UTüb, IDS, MPI are key players

all relevant metadata categories defined and entered
ISOcat already in production, API is ready
much more in ISOcat, all in collaboration with ISO TC37

Munich, October 17 2009

Metadata Organizing/Editing Tool

D-SPIN



CLARIN

but next steps are:

- support for CMDI components
- support of component/profile registry
- one of the most promising tools based on years of experience

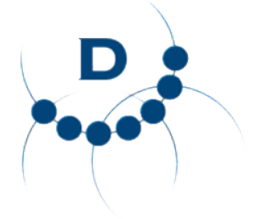
- just finished new metadata organizer/editor
- currently based on IMDI schema

Munich, October 17 2009

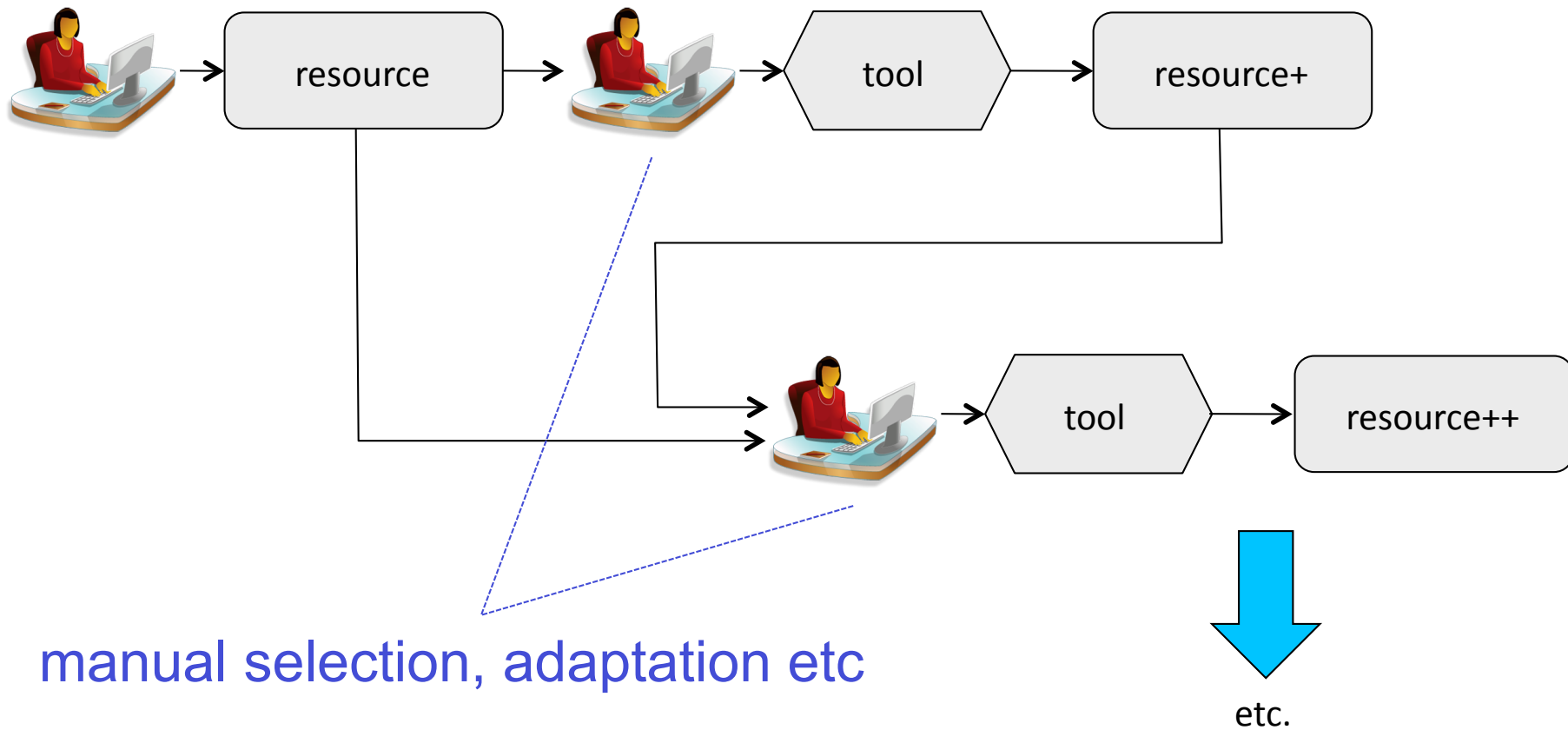
State Pillars - Generic SOA

D-SPIN

CLARIN

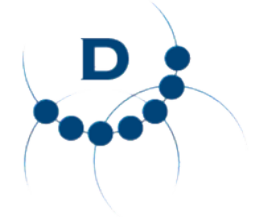


traditional workflow

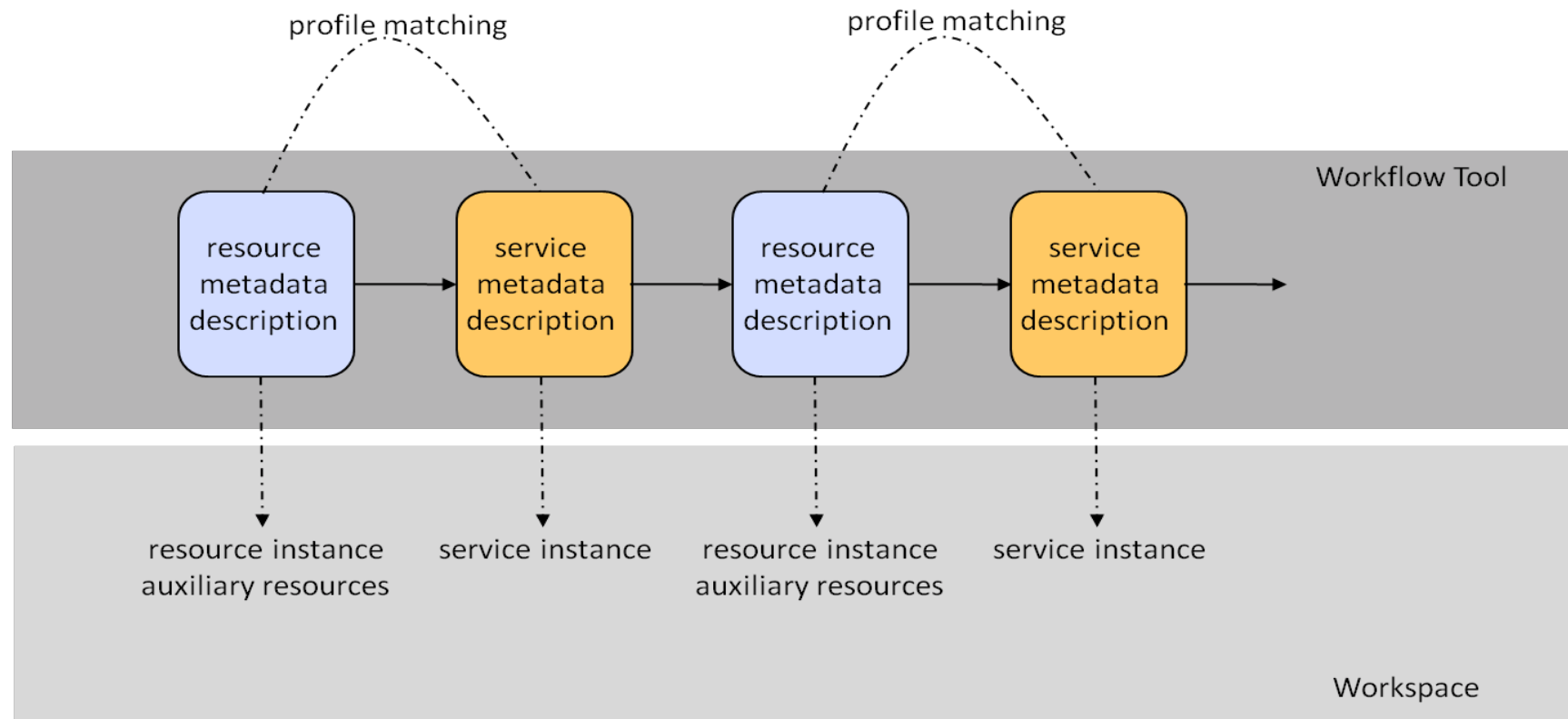


manual selection, adaptation etc

etc.



intended workflow



- metadata of resources and services are used to identify compliance
- if not they should suggest converters, etc.
- metadata specifications ready for profile matching

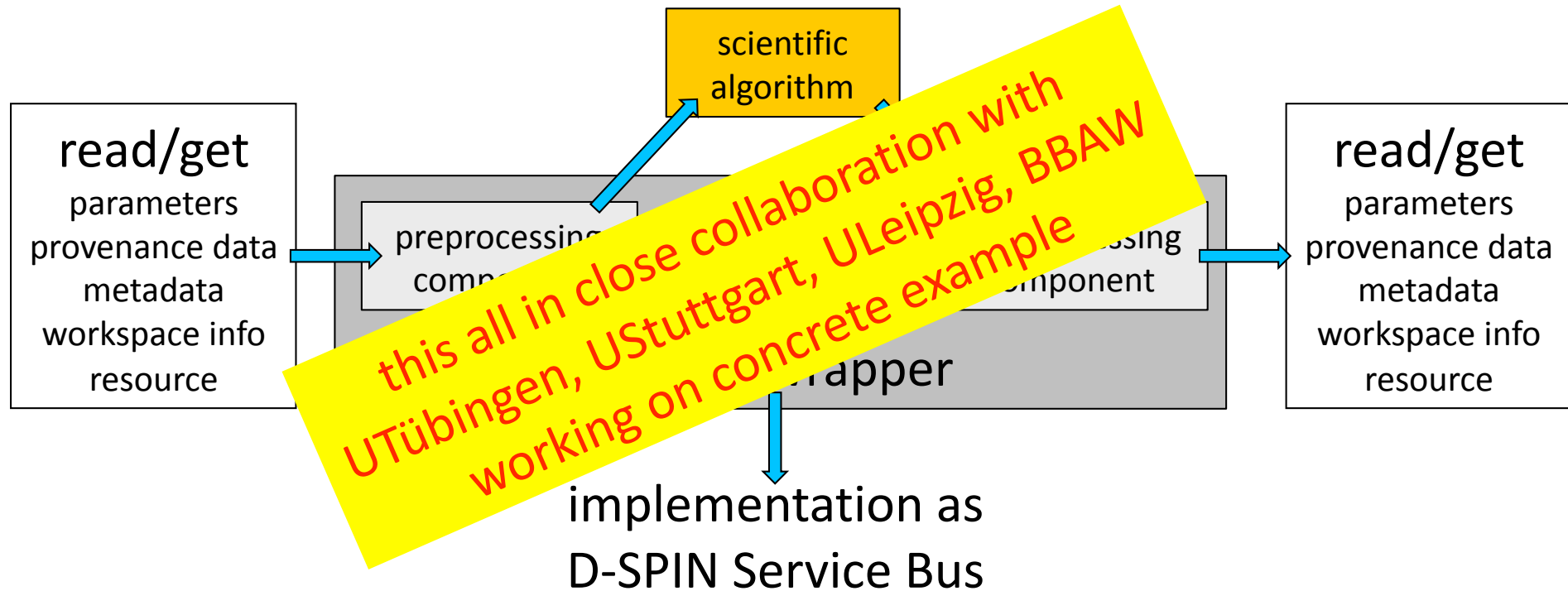
State Pillars - Generic SOA

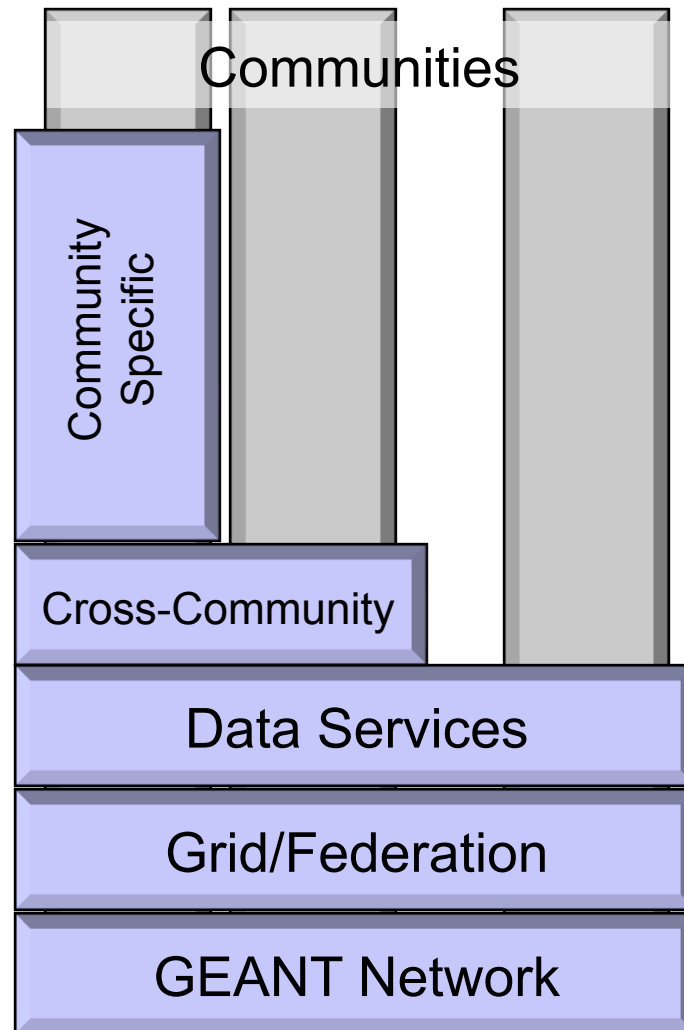
D-SPIN

CLARIN



service view





- close connection with DFN (certificates, AAI, ?)
- close connection with Grid/Federation (PID, security, monitoring, compute grid,?)
- close connection with future data services (archiving, authenticity, replication, workspaces, execution spaces, ?)
- cross-community needs to evolve (ISOcat, CMDI, SOA, ?)
- close interaction with DARIAH
- participation in TextGrid
- much in flux of course

Summary

D-SPIN

CLARIN



- building TRI in SSH costs time
 - due to complex community processes
 - need to interact at European level
 - researchers need to be convinced to rely on RI in competitive times
- still in preparatory phase, i.e. testing various components
- e-Infrastructures often not ready yet
- in general good in time to achieve goals
- but building RI is nothing for impatient people





Thanks for your attention!

